

# Trinucleotide expansion in disease: why is there a length threshold?

Do-Yup Lee<sup>1</sup> and Cynthia T McMurray<sup>2</sup>



Trinucleotide repeats (TNRs) expansion disorders are severe neurodegenerative and neuromuscular disorders that arise from inheriting a long tract (30–50 copies) of a trinucleotide unit within or near an expressed gene (Figure 1a). The mutation is referred to as ‘trinucleotide expansion’ since the number of triplet units in a mutated gene is greater than the number found in the normal gene. Expansion becomes obvious once the number of repeating units passes a critical threshold length, but what happens at the threshold to render the repeating tract unstable? Here we discuss DNA-dependent and RNA-dependent models by which a particular DNA length permits a rapid transition to an unstable state.

## Addresses

<sup>1</sup> Department of Bio and Fermentation Convergence Technology, Kookmin University, 136-702, Republic of Korea

<sup>2</sup> Life Sciences Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA

Corresponding author: McMurray, Cynthia T. ([ctmcmurray@lbl.gov](mailto:ctmcmurray@lbl.gov))

**Current Opinion in Genetics & Development** 2014, **26**:131–140

This review comes from a themed issue on **Molecular and genetic bases of disease**

Edited by **Cynthia T McMurray** and **Jan Vijg**

For a complete overview see the [Issue](#) and the [Editorial](#)

Available online 1st October 2014

<http://dx.doi.org/10.1016/j.gde.2014.07.003>

0959-437/Published by Elsevier Ltd. This is an open access article under the CC BY-NC-SA license (<http://creativecommons.org/licenses/by-nc-sa/3.0/>).

Trinucleotide expansion is the underlying basis for disease toxicity in a number of severe hereditary diseases [1,2<sup>•</sup>,3<sup>•</sup>], and occurs both in the germ line and in somatic tissues with age. The general steps of expansion in simple terms are three: structure formation, heteroduplex resolution, and gap filling synthesis (Figure 1b). Over the past years, many reviews (including our own) have focused on the first step: how heteroduplex DNA structures form [1,4<sup>•</sup>,5<sup>•</sup>,6<sup>•</sup>] (Figure 1c). Indeed, all data are consistent with a model in which heteroduplex structures are the basis for expansion, which arises broadly from classes of *de novo* excision repair, replication errors, and replication arrest and restart [1,4<sup>•</sup>,5<sup>•</sup>,6<sup>•</sup>]. All of these mechanisms invoke their own machinery to carry out heteroduplex resolution, and distinct polymerases to complete gap-filling synthesis (Table 1). DNA expansion itself appears to be independent of position of the repeat tract, other

than that it must reside in or around genes to cause observable abnormalities (Figure 1). But how do expansions begin? Here, we will consider one of the oldest questions and most puzzling feature of expansion: its length threshold.

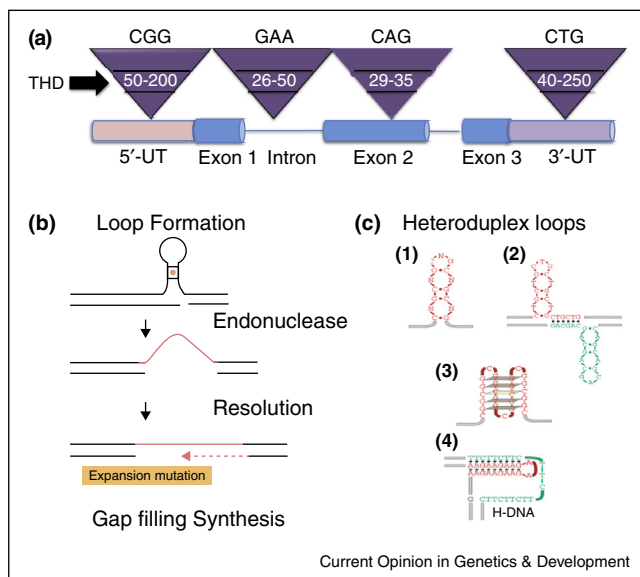
What is an expansion threshold? Expansion observed in all TNR diseases requires a pre-existing long tract of TNRs units before there is a significant probability of instability (Figure 1a). Normal allele lengths are stable, and there is no ‘jumping’ from a normal to a disease tract length [7,8] (Figure 1a). Only when an allele is of critical copy number (the threshold) does expansion become probable within the lifetime of a human, and modulate a transition from pre-mutation to full-mutation length TNR tract [7–10]. The fact that expansion becomes probable only after a threshold length is reached suggests that expansion is strongly DNA-dependent, but why does tract length matter? In this review, we discuss three major models that provide possible explanations for a length threshold in light of recent findings: firstly length-dependent reannealing of DNA or DNA–RNA hybrids, secondly coding for a minimum length of RNA and protein sufficient to induce toxicity, and finally metabolism. These mechanisms are not mutually exclusive, but some are more likely than others.

## Does the stability and size of heteroduplex loops in DNA govern the threshold?

One of the oldest and perhaps most intuitive explanation for a threshold is a minimal length at which a heteroduplex DNA intermediate becomes stable [1,2<sup>•</sup>,3<sup>•</sup>,4<sup>•</sup>,5<sup>•</sup>,6<sup>•</sup>]. Indeed, we demonstrated as early as 1995 that triplet repeats formed hairpins with repeating units of two CG pairs and a mismatch, which explained their aberrant migration on gels [11]. At the same time, Wells and co-workers observed that instability occurred in bacteria by slippage [12]. However, a structural stability model for threshold is not entirely satisfying. Loop sizes of only a few repeats are thermodynamically stable in replication slippage reactions [6], and the MutL endonuclease that resolves small loops in DNA operates efficiently at 1–4 contiguous triplet units [13].

However, the sizes of the heteroduplex loops that occur during repair are expected to be larger. The excision patch of transcription coupled repair (TCR) and nucleotide excision repair (NER) is typically around 15–20 bases [14<sup>•</sup>], corresponding to a fold-back structure of 5–7 repeats. Strand displacement during long patch BER is

Figure 1



Features of expansion and its threshold. **(a)** Generic representation of threshold limits for some representative disease alleles from distinct TNR disorders. The inverted purple triangles represent the size ranges associated with the normal, threshold, and disease length TNR alleles. In white is the threshold length for representative TNRs: CGG in the 5'-untranslated (5'UT) region characterizes the FMR-1 gene; GAA in an intron (lines) characterizes Friedreich's ataxia gene; CAG in a coding region (exon) characterizes the Huntington's gene; CTG in the 3'-untranslated (3'UT) region characterizes the myotonic dystrophy 1 (DM1) gene. The threshold limit is also referred to as the premutation length, as all full mutations arise from lengths at the upper range of normal allele and the lower edge of disease allele lengths. Below that range are stable normal repeats, and above the ranges at which expansion exists. The premutation lengths as shown are approximate sizes since there is no precise range THD is threshold. **(b)** The three most basic steps of expansion. **(c)** Distinct types of heteroduplex DNA loops are proposed as precursors to expansion: hairpins (1); cruciform (2); quadraplex (3); H-DNA triplet helix (4). (c) is taken from [5\*\*].

around the same size or larger when CAG TNRs are the repair substrate [15\*\*,16]. Moreover, small chemical lesions such as 8-oxo-guanine can trigger a switch to translesion synthesis by Pol  $\eta$  in yeast [17\*\*]. Polymerase pausing is noted in long non-coding TNRs, and the size of the loops formed during fork reversal [18] or strand-switching [19] mechanisms have the potential to promote even larger loops. The endonucleases (Table 1) that resolve the larger loops and their integration into genomic DNA are, as yet, unknown [20,21,22\*\*,23–36,37\*\*,38].

A kinetic model for the threshold on the DNA level is more likely. At any single strand break or on Okazaki fragments, free ends are in flux on and off DNA, and there is inherent competition between duplex reformation (no mutation) and structure formation at the frayed end (mutation intermediate). The threshold transition length may simply reflect the length at which the lifetime of self-pairing in

heteroduplex DNA becomes long enough to exceed the rate of gap filling synthesis (which would prevent duplex reannealing). The resulting flap folds-back to initiate structure formation at the TNR sequence. Indeed, we tested at least part of this idea by following duplex reannealing of complementary hairpins of 10 (lower than threshold) and 25 CAG repeats (at the threshold) [39]. The rate of duplex reannealing for the 25 units was one to six fold slower than the 10 units CAG repeat hairpin, although they were of similar stability. The hairpin structure of 25 units re-formed duplexes reannealed roughly 50-fold slower relative to unstructured random sequences, unstructured scrambled CAG nucleotides, and dinucleotide repeating sequences of identical length [39]. Many more constraints occur *in vivo*, and whether the lifetime of long flaps exceeds the rate of gap filling synthesis *in vitro* or *in vivo* remains to be measured. Nonetheless, the kinetic lifetime of the fold-back structure distinguishes a CAG/CTG tract at the threshold from shorter CAG/CTG tracts by the reannealing rate.

### The role of RNA–DNA hybridization in determining the threshold

But could RNA determine the DNA threshold for expansion? Reannealing kinetics appears to be relevant for a TNR threshold mechanism that is R-loop dependent [40,41]. RNA–DNA hybrids form at the expanded ( $n > 200$  rpts) but not normal CGG repeat regions (commonly 30 rpts) in the FMR1 gene from human iPSCs that were differentiated in culture for 30–60 days [40]. The majority of the RNA–DNA duplex occurs between 200 and 300 bp on either side of the expanded CGG tract, consistent with the notion that the promoter harboring the transcribed CGG-repeat tract is the binding site for the FMR1 mRNA. Transcription through the GC-rich *FMR1* 5'UTR region favors R-loop formation, with the nascent (G-rich) RNA forming a stable RNA:DNA hybrid with the template DNA strand (Figure 2a,b), thereby displacing the DNA strand. Recruitment of the TCR machinery at the stalled site may promote nicking and expansion at the site for repair during removal of the RNA–DNA hybrid block (Figure 2c). In the iPSC system, binding of the *FMR1* mRNA to the genomic repeat does not occur before day 45, implying that the hybrid forms slowly [40]. Thus, the size of a stable hybrid might determine the length at which an open transcription bubble 'sensitizes' the TNR sensitive to damage (Figure 2a) and render it subject to TCR or BER (Figure 2c). Alternatively, the RNA–DNA bubble may be the threshold 'impediment' needed for 'calling in' fork reversal [18] or strand-switching [19] resolution mechanisms.

Because of patient variability, it is difficult to determine the precise relationship among transcriptional silencing, the size of the RNA–DNA hybrid, or the level of chemically modified bases. Missing from the iPSC experiments are robust measures of the DNA methylation status and

**Table 1****Mammalian structure-specific endonucleases**

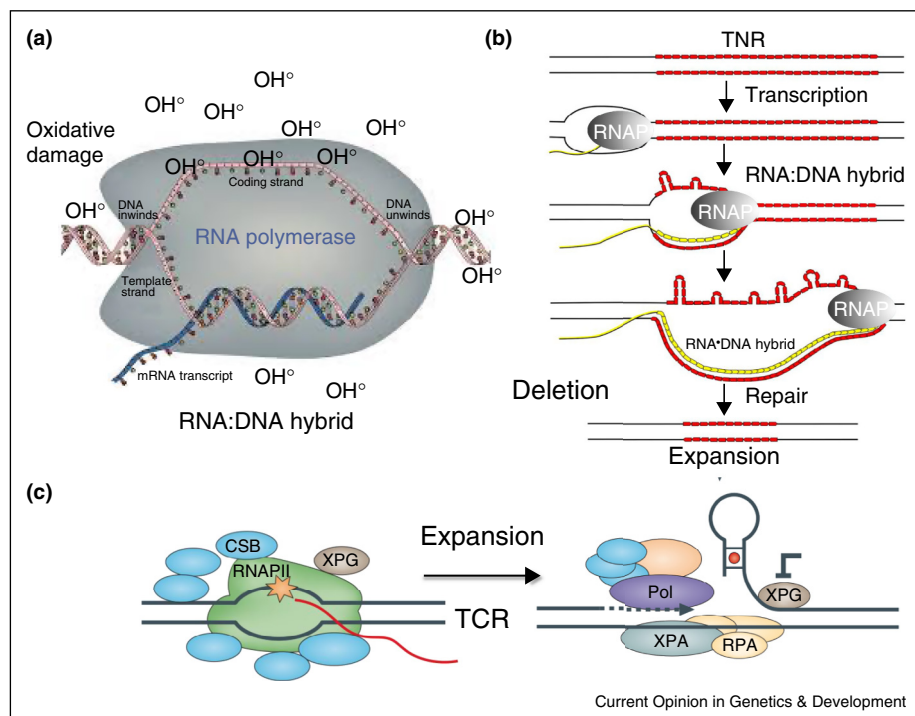
Molecule	Loop size ( <i>n</i> = # rpts)	Pathway	Refs	Endonuclease function
MutL	<i>n</i> < 4 rpts	MMR	[21]	Removal of mismatches and small loops
MRE11	<i>n.d.</i>	DSBR	[22**,23]	3'-5' endonuclease activity at DSBs; hairpin cleaving activity
MRE11/Sae2				Hairpin cleaving activity with a single strand nick
APE	<i>n</i> > 4 rpts	BER	[24]	Cleaves 5' side of abasic site generating a 3'-OH group for polymerase extension
XPF/ERCC1	<i>n</i> > 4 rpts	NER/TCR	[25,26]	Cleaves a 5' bubble substrates
XPG	<i>n</i> > 4 rpts	NER/TCR	[26]	Cleaves a 3' bubble substrates
Mus81/Eme1	<i>n.d.</i>	DSBR	[27]	Structure-specific 3'-flap DNA endonuclease that can process substrates resembling replication intermediates
SLX	<i>n.d.</i>	DSBR	[28,29]	Endonuclease activity toward replication forks, 5'-flaps, and Holliday junctions
FEN-1		Replication/gap filling repair	[30]	Member of the XPG/RAD2 endonuclease family. Removes 5' overhanging flaps in DNA repair; processes the 5' ends of Okazaki fragments in lagging strand DNA synthesis; cofactor in gap filling repair-dependent replication DNA secondary structure can inhibit FEN-1 flap processing
Gen1/YEN1	<i>n.d.</i>	DSBR, HR	[31]	Holliday junction 5' flap endonuclease and resolves Holliday junctions
ZRANB3	<i>n.d.</i>	Replication stress	[32,33]	Localizes to DNA replication sites and interacts with the components of the replication machinery. Maintains genome stability at stalled or collapsed replication forks by facilitating fork restart and limiting inappropriate recombination that could occur during template switching events; acts as a structure-specific endonuclease that cleaves the replication fork D-loop intermediate, generating an accessible 3'-OH group in the template of the leading strand. Cleaves branched DNA structures with unusual polarity
GQN1	<i>n.d.</i>	Somatic hypermutability	[34]	(G quartet nuclease 1) cleaves within the single-stranded region 5' of the barrel formed by stacked G quartets. GQN1 does not cleave duplex or single-stranded DNA, Holliday junctions, or G4 RNA
SNM1	<i>n.d.</i>	ICL	[35]	Operates predominantly in interstrand crosslink (ICL) repair; structure-specific DNA hairpin opening endonuclease
RNase H	<i>n.d.</i>	RNA/DNA duplexes	[36]	RNase H is a ribonuclease that cleaves the RNA in a DNA/RNA duplex to produce ssDNA. RNase H is a non-specific endonuclease and catalyzes the cleavage of RNA via a hydrolytic mechanism, aided by an enzyme-bound divalent metal ion. RNase H leaves a 5'-phosphorylated product
Dicer	RNA	Double stranded RNA	[37**]	Dicer is an endoribonuclease in the RNase III family that cleaves double-stranded RNA (dsRNA)
Cas9 breaks	RNA	DNA/RNA DNA editing	[38]	RNA-guided DNA endonuclease to generate double-strand

alterations of the CGG tract length that might have occurred during a 30–60 day differentiation period [40]. Extensive methylation in the promoter region at CGG repeats accompanies transcriptional suppression [42]. If the RNA–DNA hybrid triggers methylation and heterochromatin formation, then another attractive model for expansion is the removal of methylated bases and DNA loop formation via BER [43]. Although removal of methylated bases by BER is accomplished by several DNA glycosylases with different specificities, none are known to promote TNR expansion. In fact, expansion is likely to occur in unmethylated state: (1) Rare individuals having full mutations but normal intelligence lack hypermethylation and maintain expression of *FMR1* mRNA [44]. (2) Pharmacologic treatment with the DNA methylation

inhibitor 5-aza-2'-deoxycytidine (azadC) reactivates transcription and FMRP expression but does not alter the repeat tract [45]. More likely, expansion is a response to the stress of RNA-induced or protein-induced aggregation/toxicity, which enhances oxidative damage in DNA. Removal of the oxidized bases by the BER or TCR pathways results in loop formation and expansion. Indeed, loss of OGG1 [15\*\*], NEILS 1 [46], and XPA [47] reduces expansion in mice. Novel mechanisms for enhancing oxidative damage and toxicity are discussed below.

Whether RNA–DNA hybrids form at TNRs in other non-coding regions (which generate large expansions) is unknown. In coding regions, the expanded CAG/CTG

Figure 2



**(a)** A schematic representation of an RNA–DNA hybrid and its role in hypothetical mechanisms in TNR expansion during transcription. (a) In the FMR1 gene, the RNA–DNA hybrid is proposed to form at the promoter region including the 5′ untranslated region harboring CGG repeat. The paused RNA polymerase stabilizes the open transcription bubble, which is susceptible to oxidative DNA damage. A possible model for expansion is oxidative stress and base damage which is removed during BER [2<sup>°</sup>,15<sup>°</sup>]. **(b)** A possible model for TNR deletion at the RNA–DNA hybrid site. The open non-transcribed strand forms self-paired DNA loops along the transcription bubble as polymerase passes [2<sup>°</sup>,5<sup>°</sup>]. **(c)** TCR may also play a role by removing the RNA–DNA hybrid. At or below the threshold, the stalled polymerase may successfully recruit the TCR machinery or utilize it in some way to remove the RNA–DNA hybrid. If XPG is inhibited, a flap structure folds-back to form a structural intermediate for expansion. These loops are most likely to result in deletion after subsequent removal, although expansion is possible if an endonuclease is able to clip the hairpin loops on the unpaired DNA strand, and the TNR duplex is restored by gap filling synthesis [2<sup>°</sup>,5<sup>°</sup>].

repeat tracts ( $n > 35$  rpts) overlap in length with those of the FMR-1 ‘normal’ CGG range [1,2<sup>°</sup>,3<sup>°</sup>,4<sup>°</sup>,5<sup>°</sup>,6<sup>°</sup>] (commonly 30 rpts), which does not form hybrids. Moreover, CAG expansions do not impose transcription silencing of their respective genes [1,3]. If a minimum DNA–RNA hybrid causes the transcriptional silencing at a threshold length, then it is unlikely to be a mechanism that is common to all TNR genes.

### Double stranded RNA models for expansion.

Another consideration in a RNA-dependent hybridization model for threshold is the effect, if any, of bi-directional transcription of the TNR region [48<sup>°</sup>]. For example, several novel anti-sense FRM1 transcripts exist in the FRM1 locus (ASFMR4-6), and some overlap the CGG repeat region [49]. ASFMR4 transcript is spliced, polyadenylated and exported to the cytoplasm [42,49]. If a bi-directional transcript overlaps with the sense transcript, double stranded RNA is formed as a Dicer substrate. It is not easy to imagine how short siRNA hybrids within the

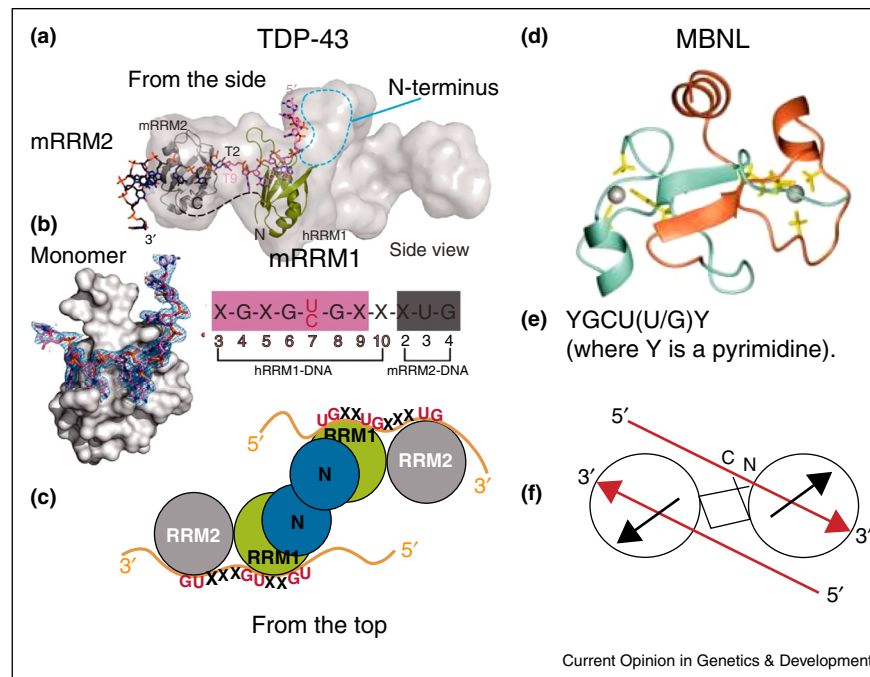
TNR tract results directly in expansion. Either multiple siRNA binding creates a RNA–DNA hybrid of similar length to that of an mRNA hybrids [40], and are removed by similar mechanisms, or the shorter RNA–DNA hybrid opens the DNA sufficiently to increase exposure to oxidative DNA damage at a preferred threshold length (Figure 2a).

### Protein–RNA interaction models for the threshold

New models provide insight on how RNA–protein complexes of threshold length might provoke chemical lesions in DNA, and lead to expansion. TAR-DNA-binding protein 43 (TDP-43) [50] is poised to bind to a RNA–DNA hybrid. TDP-43 is a dimeric protein with two RNA recognition motif (RRM) domains that bind both DNA and RNA [50,51<sup>°</sup>,52] (Figure 3a–c), and interact with fragile X mental retardation protein (FMRP) in an (FMRP)/Staufen (STAU1) complex [53]. This complex forms aggregates analogous to those of polyglutamine



Figure 3



The structural models of TDP-43 and MBNL bound to nucleic acids are similar. **(a)** TDP-43 binds to both DNA and to RNA. The small angle X-ray scattering (SAXS) envelope of TDP-43 dimer is fitted with the crystal structure of RNA recognition motif (hRRM1)-DNA and mRRM2-DNA in the orientation that the DNA forms a continuous 5'–3' strand, as it is bound from hRRM1 to mRRM2 in TDP-43. **(b)** SAX structure of a TDP-43 monomer bound to ssDNA. The putative RNA binding sequence of TDP-43 is derived from the sequence of two tandem RNA recognition motifs (RRM), hRRM1-DNA and mRRM2-DNA complexes. The RRM1 binds to RNA with extensive contacts with the conserved  $\beta$ -sheet residues and loop residues. **(c)** Schematic representation of the domains structure of TDP-43 homodimer, which binds to a long UG-rich RNA via its RRM1 and RRM2 domains. **(d)** Ribbon diagram presentation of zinc finger domain of MBNL2, with the side chains of the ligand residues and zinc ions (Gray balls) and binds to a GU-rich sequence **(e)**. **(f)** Like TDP-43, MBNL2 is also capable of forming a dimer. Each circle represents a CCH-type zinc finger motif, and the black arrows represent the zinc finger orientation corresponding to the direction of the first helix in each of the domains. The red arrows indicate the 5'–3' direction of the two accommodated RNA on the TZF12 of the MBNL2 dimer.

proteins, which induce cellular stress and oxidative DNA damage. The DNA length at which the encoded RNA forms aberrant protein–RNA complexes may be the threshold for the enhanced stress.

The mechanisms of RNA aggregate formation are unknown, but it is likely due to the disruption of complex formation at its C-terminus. TDP-43 interacts at its C-terminus with the hnRNP family of translation factors, as well as the splicing factors muscleblind (MBNL) and CUG-BP1 (CUG binding protein 1) [54]. MBNL and CUG-BP1 impart two opposing effects on splicing, and they occur through binding of distinct regions of the target RNA [55]. Both CUG-BP1 and MBNL bind to short-structured CUG and CCUG repeats in RNA with high affinity and specificity [55] (Figure 3d,e). Only 6 base pairs are necessary for MBNL binding: two pyrimidine mismatches and four guanosine–cytosine base pairs that form in a helical region of a stem-loop in the endogenous pre-mRNA target [55] (Figure 3e). In the myotonic

dystrophy gene (DM1), these two regions of the RNA reside on the 3' and 5' sides that surround the TNR [56]. The length of the TNR tract affects only MBNL binding and impairs its function. A loss-of-function in MBNL and a gain-of-function in CELF4 tend to favor generation of the alternatively spliced forms.

TDP-43 also binds to both the 3' and 5' end of the DM1 mRNA, and raises the possibility of that binding of MBNL and TDP-43 occurs at the same sites. Whether these two proteins overlap in the recognition to mRNA is unknown, but the common binding sites and functionality in the DM1 mRNA raise the possibility that the bi-partite mRNA binding at the C-terminus of TDP-43 integrates translation and splicing activity. Interestingly, TDP-43 controls its own expression through a negative feedback loop involving interactions with its mRNA at the 3' end [57]. Furthermore, the domain structure of TDP-43 is similar to that of both heterogeneous nuclear ribonucleoprotein (hnRNP) and muscleblind (MBNL)

[58] (Figure 3f): an N-terminal domain (NTD) and two tandem RNA recognition motifs (RRM1 and RRM2), followed by a C-terminal glycine-rich region (G) (Figure 3a–c). The C-terminus of TDP-43 acts as a hub that regulates both splicing and translation. Indeed, TNR coding transcripts are associated with an unusual type of translation, Repeat Associated Non-ATG translation (RAN-translation) [59\*\*]. RAN-translation does not require an ATG translation start site, and random translation at TNRs occurs in all reading frames [59\*\*].

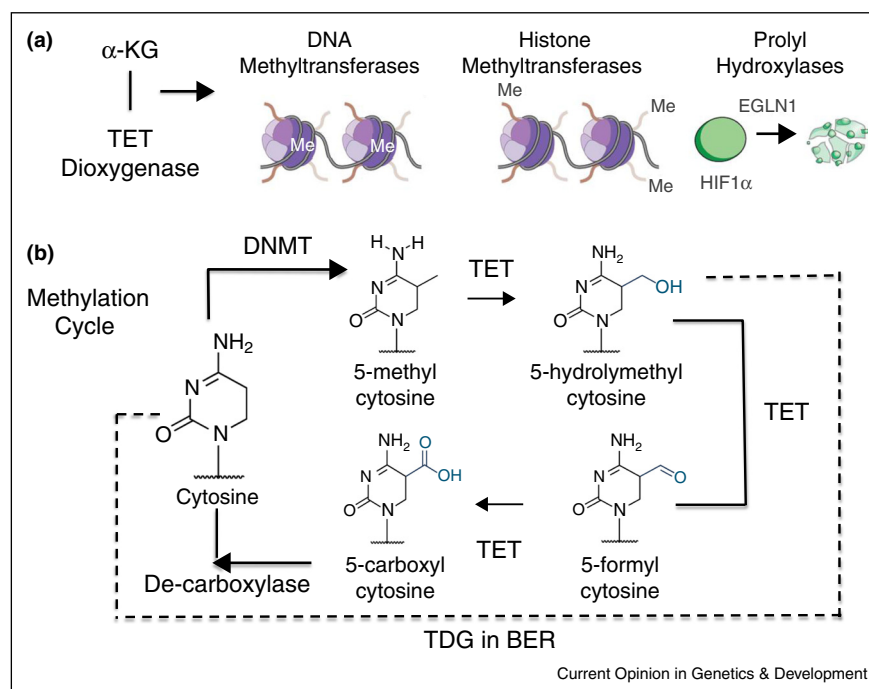
Given its hub-like features, maintaining the C-terminus of TDP-43 would appear to be a key regulatory process. Indeed, pathological TDP-43 in the cytoplasmic and intranuclear inclusions is hyper-phosphorylated, ubiquitinated, and cleaved to ~25 kDa C-terminal fragments in affected brain regions [60]. C-terminal-deleted TDP-43 without the glycine-rich tail is sufficient to form a head-to-head homodimer primarily via its N-terminal domain, which form fibrils *in vitro* [60]. Thus, proteolytic cleavage of TDP-43 within the RRM2 removes the N-terminal dimerization domain and produces unassembled truncated RRM2 fragments, which can abnormally oligomerize into

high-order inclusions (Figure 3). The resulting increase in oxidative DNA damage promotes expansion indirectly by RNA-mediated depletion of TDP-43/FMRP/STAU1 in the nucleus and an increase in cellular stress.

### The role of maintaining balanced DNA methylation in expansion

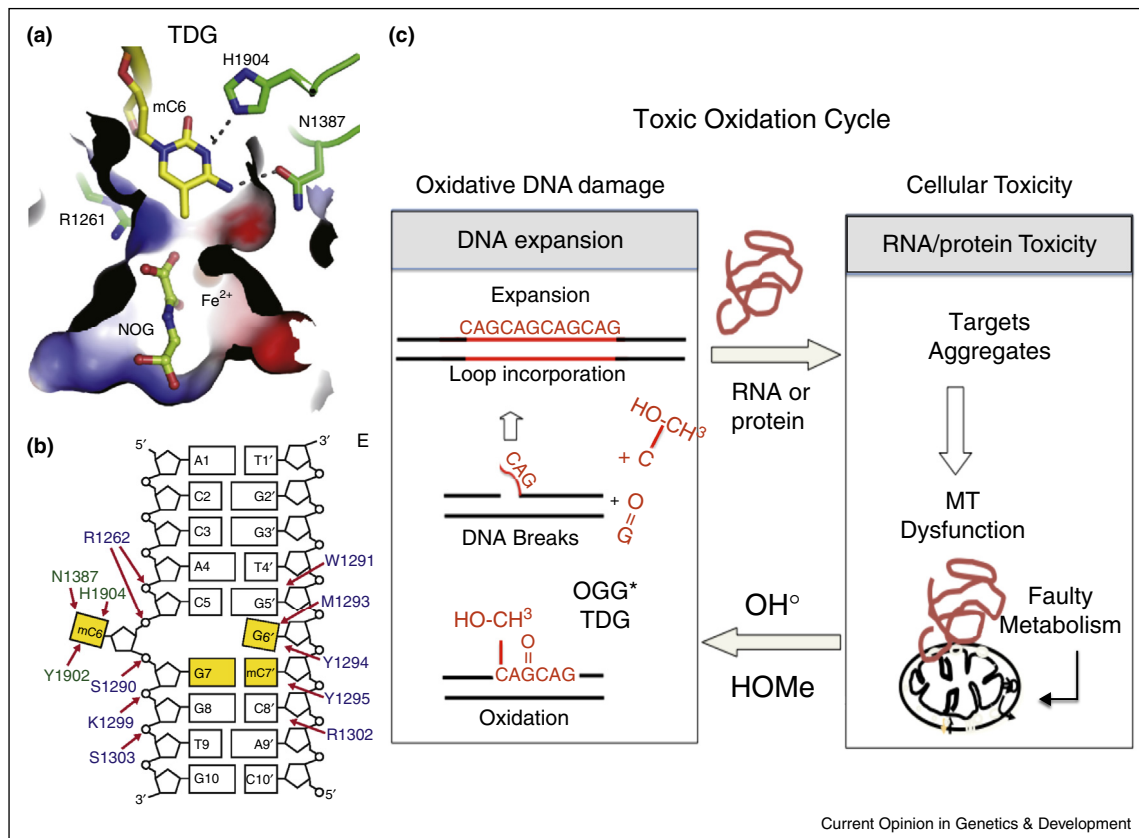
Whether this type of RNA-mediated mechanism applies to all triplet repeat disorders is unknown, but there are direct links between them and mitochondrial metabolism. A TDP-43 binds to the mRNA of the silent information regulator 1 (SIRT1), which is implicated in double-stranded DNA break repair and DNA metabolism in all cells [61]. SIRT1 is an NAD<sup>+</sup> dependent class III histone deacetylase [61], which cooperates with elongation factor 1 (E2F1) to regulate apoptotic response to DNA damage. SIRT1 knockdown results in poly Q-expanded aggregation of androgen receptor (AR) and  $\alpha$ -synuclein [62], consistent with a role of the SIRT1mRNA-TDP-43 complex in aggregation, and supports the notion that RNA processing by TDP-43 and chromatin organization SIRT1 are functionally connected. TDP-43 regulates alternate splicing of the CFTR

Figure 4



The methylation cycle and its metabolic regulation. (a)  $\alpha$ -Ketoglutarate ( $\alpha$  KG) is a critical co-factor for the ten-eleven translocation dioxygenase (TET) which oxidizes 5-methylcytosine. TET regulates both DNA and histone methylation, and *EGLN1* gene product, hypoxia-inducible factor prolyl hydroxylase 2 (HIF-PH2), or prolyl hydroxylase domain-containing protein 2 (PHD2), is an enzyme encoded by the *EGLN1* gene. It is also known as Egl nine homolog 1. The *EGLN1* gene is responsible for ubiquitin-proteasome degradation pathway by hydroxylation of proline-564 and proline-402 by PHD2. (b) The methylation cycle regulated by TET. The action of DNA demethyltransferases (DNMTs) converts cytosine (C) for 5' methyl C. The TET enzymes oxidize 5' methylC to 5'-hydroxylmethylC and 5'-formylC. The 5'-hydroxylmethylC is removed by the BER enzyme, TDG, to restore C, which is subsequently methylated. 5'-formylC is caboxylated by an unknown enzyme to restore C, which is subsequently methylated. Success of this cycle maintains a balance of methylation in the genome. The dotted line indicates the removal of 5-hydroxylmethylcytosine and restoration to cytosine to reset the cycle.

Figure 5



Two-state threshold model. **(a)** The binding pocket of the OH-CH<sub>3</sub>-cytosine in TDG. **(b)** The overall cartoon structure of the OH-CH<sub>3</sub>-cytosine-TDG complex. **(c)** The toxic oxidation cycle. We propose a two-state model for expansion. Expansion arises from toxicity imparted from RNA and protein-mediated toxicity by two mechanisms. RNA and protein-mediated toxicity induces mitochondrial stress and a concomitant increase in oxidative damage to DNA. The oxidative damage is removed by DNA glycosylases. 8-oxo-guanine glycosylase (OGG1)<sup>\*</sup> is a major enzyme that removes oxidative damage, but it can also be removed by the NEILS 1<sup>\*</sup> glycosylase, or the machinery of TCR<sup>\*</sup>. Single strand break intermediates arise during base removal and produce flaps that fold-back to generate structural intermediates for expansion.

RNA at the intron8/exon9 junction, implying that alternative splicing may have a direct consequence on the chromatin organization, which is altered at long, congenital TNR lengths.

Interestingly, isocitrate dehydrogenase 1 (IDH1) and IDH2 catalyze the interconversion of isocitrate and  $\alpha$ -ketoglutarate ( $\alpha$ -KG) [63] (Figure 4a).  $\alpha$ -KG is a TCA cycle intermediate in mitochondria, and is an essential co-factor for many enzymes, including JmJc domain-containing histone demethylases [63,64<sup>••</sup>], and a family of 5-methylcytosine (5mC) hydroxylases, Ten-eleven translocation dioxygenase (TET) [64<sup>••</sup>] and EglN prolyl-4-hydroxylases (Figure 4a). Both TET1 and TET3 proteins contain a DNA-binding motif that is believed to target CpG sites (Figure 4a). TET2 converts 5-methylcytosine to 5-hydroxymethylcytosine (5-hmC) in DNA and uses  $\alpha$ -ketoglutarate as a co-substrate [65]. The resulting (5-hmC) is removed by the BER enzyme

thymine DNA glycosylase (TDG) [64<sup>••</sup>] (Figure 4b). At the excision site, cytosine replaces 5-hmC, and methylation occurs subsequently to restore the methylated state and 5-mC [64<sup>••</sup>] (Figures 4b and 5a,b). Thus, metabolism is apparently a regulatory mechanism to maintain a balanced methylation state, and influences expansion. Since methylation status does not appear to play a role in expansion *per se*, RNA-induced and protein-induced toxicity may act in a feed-back loop, producing a toxic oxidation cycle and expansion during removal of the oxidative DNA damage (Figure 5c).

## Conclusions

Although new possibilities for DNA-mediated, RNA-mediated and protein-mediated toxicity are emerging, these diverse pathways, in the end, are likely to induce expansion by similar mechanisms (Figure 5). Physically, expansion occurs by loop formation at free DNA ends during DNA excision, by polymerase slippage or by

strand switching events that occur during replication or fork-reversal. From this simple viewpoint, we can construct both physical and functional definitions of an expansion threshold. Physically, the threshold defines a kinetic point in which self-pairing ‘wins’ over duplex reformation. Structures form at Okazaki fragment ends and/or at single strand breaks are trapped by gap filling synthesis or continued replication (Figure 5). Functionally, the threshold is likely to be the limiting length at which lesion load induces DNA repair.

## Acknowledgement

This work was supported by National Institutes of Health grants ES020766-01 (to CTM), and CA092584 (CTM), and NS060115 (CTM).

## References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Nelson DL, Orr HT, Warren ST: **The unstable repeats — three evolving faces of neurological disease.** *Neuron* 2013, **77**:825-843.
2. McMurray CT: **Mechanisms of trinucleotide repeat instability during human development.** *Nat Rev Genet* 2010, **11**:786-799.  
A comprehensive review of recent progress in linking the features of human disease with the replication and repair-dependent mechanisms of expansion and how they are used during different stages of development.
3. La Spada AR, Taylor JP: **Repeat expansion disease: progress and puzzles in disease pathogenesis.** *Nat Rev Genet* 2010, **11**:247-258.  
A comprehensive review of recent progress in understanding the pathophysiology of expansion disease.
4. Dion V, Wilson JH: **Instability and chromatin structure of expanded trinucleotide repeats.** *Trends Genet* 2009, **25**:288-297.  
A comprehensive review about the chromatin aspects of triplet expansion.
5. Mirkin SM: **Expandable DNA repeats and human disease.** *Nature* 2007, **447**:932-940.  
An excellent review about the replication and repair-dependent mechanisms of triplet expansion.
6. McMurray CT: **Mechanisms of DNA expansion.** *Chromosoma* 1995, **104**:2-13.  
An excellent review on the biophysical properties of triplet repeats.
7. Goldberg YP, McMurray CT, Zeisler J, Almqvist E, Sillence D, Richards F, Gacy AM, Buchanan J, Telenius H, Hayden MR: **Increased instability of intermediate alleles in families with sporadic Huntington disease compared to similar sized intermediate alleles in the general population.** *Hum Mol Genet* 1995, **4**:1911-1918.
8. Squitieri F, Jankovic J: **Huntington's disease: how intermediate are intermediate repeat lengths?** *Mov Disord* 2012, **27**:1714-1717.
9. Hoem G, Raske CR, Garcia-Arocena D, Tassone F, Sanchez E, Ludwig AL, Iwahashi CK, Kumar M, Yang JE, Hagerman PJ: **CGG-repeat length threshold for FMR1 RNA pathogenesis in a cellular model for FXTAS.** *Hum Mol Genet* 2011, **20**:2161-2170.
10. Semaka A, Kay C, Doty C, Collins JA, Bijlsma EK, Richards F, Goldberg YP, Hayden MR: **CAG size-specific risk estimates for intermediate allele repeat instability in Huntington disease.** *J Med Genet* 2013, **50**:696-703.
11. Gacy AM, Goellner G, Juranić N, Macura S, McMurray CT: **Trinucleotide repeats that expand in human disease form hairpin structures in vitro.** *Cell* 1995, **81**:533-540.
12. Kang S, Jaworski A, Ohshima K, Wells RD: **Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in E. coli.** *Nat Genet* 1995, **10**:213-218.
13. Pluciennik A, Burdett V, Baitinger C, Iyer RR, Shi K, Modrich P: **Extrahelical (CAG)/(CTG) triplet repeat elements support proliferating cell nuclear antigen loading and MutLα endonuclease activation.** *Proc Natl Acad Sci U S A* 2013, **110**:12277-12282.
14. Hanawalt PC, Spivak G: **Transcription-coupled DNA repair: two decades of progress and surprises.** *Nat Rev Mol Cell Biol* 2008, **9**:958-970.  
Excellent review on TCR.
15. Kovtun IV, Liu Y, Bjoras M, Klungland A, Wilson SH, McMurray CT: **OGG1 initiates age-dependent CAG trinucleotide expansion in somatic cells.** *Nature* 2007, **447**:447-452.  
First report linking oxidative DNA damage directly to a mechanism expansion via a base excision repair.
16. Liu Y, Wilson SH: **DNA base excision repair: a mechanism of trinucleotide repeat expansion.** *Trends Biochem Sci* 2012, **37**:162-172.
17. Rodriguez GP, Song JB, Crouse GF: **In vivo bypass of 8-oxodG.** *PLoS Genet* 2013, **9**:e1003682.  
Novel model for a small chemical lesion, 8-oxo-G, to lesion by-pass.
18. Follonier C, Oehler J, Herrador R, Lopes M: **Friedreich's ataxia-associated GAA repeats induce replication-fork reversal and unusual molecular junctions.** *Nat Struct Mol Biol* 2013, **20**:486-494.
19. Mirkin EV, Mirkin SM: **To switch or not to switch: at the origin of repeat expansion disease.** *Mol Cell* 2014, **53**:1-3.
20. Sommer D, Stith CM, Burgers PM, Lahue RS: **Partial reconstitution of DNA large loop repair with purified proteins from *Saccharomyces cerevisiae*.** *Nucleic Acids Res* 2008, **36**:4699-4707.
21. Gueneau E, Dherin C, Legrand P, Tellier-Lebegue C, Gilquin B, Bonnesoeur P, Londino F, Quemener C, Le Du MH, Márquez JA, Moutiez M, Gondry M, Boiteux S, Charbonnier JB: **Structure of the MutLα C-terminal domain reveals how Mlh1 contributes to Pms1 endonuclease site.** *Nat Struct Mol Biol* 2013, **20**:461-468.
22. Majka J, Alford B, Ausio J, Finn RM, McMurray CT: **ATP hydrolysis by RAD50 protein switches MRE11 enzyme from endonuclease to exonuclease.** *J Biol Chem* 2012, **287**:2328-2341.  
First report demonstrating the mechanism of endonuclease switching of MRE11/Rad50.
23. Westmoreland JW, Resnick MA: **Coincident resection at both ends of random, γ-induced double-strand breaks requires MRX (MRN), Sae2 (Ctp1), and Mre11-nuclease.** *PLoS Genet* 2013, **9**:e1003420.
24. Mazouzi A, Vigouroux A, Aikeshv B, Brooks PJ, Saparbaev MK, Morera S, Ishchenko AA: **Insight into mechanisms of 3'-5' exonuclease activity and removal of bulky 8,5'-cyclopurine adducts by apurinic/apyrimidinic endonucleases.** *Proc Natl Acad Sci U S A* 2013, **110**:E3071-E3080.
25. O'Neil NJ, Martin JS, Youds JL, Ward JD, Petalcorin MI, Rose AM, Boulton SJ: **Joint molecule resolution requires the redundant activities of MUS-81 and XPF-1 during *Caenorhabditis elegans* meiosis.** *PLoS Genet* 2013, **9**:e1003582.
26. Tsutakawa SE, Lafrance-Vanasse J, Tainer JA: **The cutting edges in DNA repair, licensing, and fidelity: DNA and RNA repair nucleases sculpt DNA to measure twice, cut once.** *DNA Repair (Amst)* 2014. pii:S1568-7864(14)00093-7.
27. Pepe A, West SC: **Substrate specificity of the MUS81-EME2 structure selective endonuclease.** *Nucleic Acids Res* 2014, **42**:3833-3845.
28. Wyatt HD, Sarbajna S, Matos J, West SC: **Coordinated actions of SLX1-SLX4 and MUS81-EME1 for Holliday junction resolution in human cells.** *Mol Cell* 2013, **52**:234-247.



29. Agostinho A, Meier B, Sonnevile R, Jagut M, Woglar A, Blow J, Jantsch V, Gartner A: **Combinatorial regulation of meiotic Holliday junction resolution in *C. elegans* by HIM-6 (BLM) helicase, SLX-4, and the SLX-1, MUS-81 and XPF-1 nucleases.** *PLoS Genet* 2013, **9**:e1003591.
  30. Hosfield DJ, Mol CD, Shen B, Tainer JA: **Structure of the DNA repair and replication endonuclease and exonuclease FEN-1: coupling DNA and PCNA binding to FEN-1 activity.** *Cell* 1998, **95**:135-146.
  31. Gao M, Rendtlew Danielsen J, Wei LZ, Zhou DP, Xu Q, Li MM, Wang ZQ, Tong WM, Yang YG: **A novel role of human Holliday junction resolvase GEN1 in the maintenance of centrosome integrity.** *PLoS ONE* 2012, **7**:e49687.
  32. Zakharyevich K, Tang S, Ma Y, Hunter N: **Delineation of joint molecule resolution pathways in meiosis identifies a crossover-specific resolvase.** *Cell* 2012, **149**:334-347.
  33. Weston R, Peeters H, Ahel D: **ZRANB3 is a structure-specific ATP-dependent endonuclease involved in replication stress response.** *Genes Dev* 2012, **26**:1558-1572.
  34. Sun H, Yabuki A, Maizels N: **A human nuclease specific for G4 DNA.** *Proc Natl Acad Sci U S A* 2001, **98**:12444-12449.
  35. Tiefenbac T, Junop M: **Pso2 (SNM1) is a DNA structure-specific endonuclease.** *Nucleic Acids Res* 2012, **40**:2131-2139.
  36. Yuan J, Ghosal G, Chen J: **Crystal structure of metagenome-derived LC11-RNase H1 in complex with RNA/DNA hybrid.** *J Struct Biol* 2013, **182**:144-154.
  37. Tian Y, Simanshu DK, Ma JB, Park JE, Heo I, Kim VN, Patel DJ: **A phosphate-binding pocket within the platform-PAZ-connector helix cassette of human Dicer.** *Mol Cell* 2014, **53**:606-616.
- Identification of a key gene RNA-guided editing endonuclease.
38. Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, Anders C, Hauer M, Zhou K, Lin S, Kaplan M, Iavarone AT, Charpentier E, Nogales E, Doudna JA: **Structures of Cas9 endonucleases reveal RNA-mediated conformational activation.** *Science* 2014, **343**:1247997.
  39. Gacy M, McMurray CT: **Influence of hairpins on template reannealing at trinucleotide repeat duplexes: a model for slipped DNA.** *Biochemistry* 1998, **37**:9426-9434.
  40. Colak D, Zaninovic N, Cohen MS, Rosenwaks Z, Yang WY, Gerhardt J, Disney MD, Jaffrey SR: **Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome.** *Science* 2014, **343**:1002-1005.
  41. Loomis EW, Sanz LA, Chédin F, Hagerman PJ: **Transcription-associated R-loop formation across the human FMR1 CGG-repeat region.** *PLoS Genet* 2014, **10**:e1004294.
  42. Ladd PD, Smith LE, Rabaia NA, Moore JM, Georges SA, Hansen RS, Hagerman RJ, Tassone F, Tapscott SJ, Filippova GN: **An antisense transcript spanning the CGG repeat region of FMR1 is upregulated in premutation carriers but silenced in full mutation individuals.** *Hum Mol Genet* 2007, **16**:3174-3187.
  43. Lai Y, Beaver JM, Lorente K, Melo J, Ramjagsingh S, AgoulNIK IU, Zhang Z, Liu Y: **Base excision repair of chemotherapeutically-induced alkylated DNA damage predominantly causes contractions of expanded GAA repeats associated with Friedreich's ataxia.** *PLOS ONE* 2014, **9**:e93464.
  44. Hagerman RJ, Hull CE, Safanda JF, Carpenter I, Staley LW, O'Connor RA, Seydel C, Mazzocco MMM, Snow K, Thibodeau SN: **High functioning fragile X males: demonstration of an unmethylated fully expanded FMR-1 mutation associated with protein expression.** *Am J Med Genet* 1994, **51**:298-308.
  45. Pietrobono R, Pomponi MG, Tabolacci E, Oostra B, Chiurazzi P, Neri G: **Quantitative analysis of DNA demethylation and transcriptional reactivation of the FMR1 gene in fragile X cells treated with 5-azadeoxycytidine.** *Nucleic Acids Res* 2002, **30**:3278-3285.
  46. Møllersøn L, Rowe AD, Illuzzi JL, Hildrestrand GA, Gerhold KJ, Tvetters L, Bjølgerud A, Wilson DM 3rd, Bjørås M, Klungland A: **Neil1 is a genetic modifier of somatic and germline CAG trinucleotide repeat instability in R6/1 mice.** *Hum Mol Genet* 2012, **21**:4939-4947.
  47. Hubert L Jr, Lin Y, Dion V, Wilson JH: **Xpa deficiency reduces CAG trinucleotide repeat instability in neuronal tissues in a mouse model of SCA1.** *Hum Mol Genet* 2011, **20**:4822-4830.
  48. Batra R, Charizanis K, Swanson MS: **Partners in crime: bidirectional transcription in unstable microsatellite disease.** *Hum Mol Genet* 2010, **19**:R77-R82.
- Excellent summary of the discovery and important cooperative activities of muscleblind and CUG bindin protein.
49. Pastori C, Peschansky VJ, Barbooth D, Mehta A, Silva JP, Wahlestedt C: **Comprehensive analysis of the transcriptional landscape of the human FMR1 gene reveals two new long noncoding RNAs differentially expressed in fragile X syndrome and fragile X-associated tremor/ataxia syndrome.** *Hum Genet* 2014, **133**:59-67.
  50. Tollervey JR, Curk T, Rogelj B, Briesse M, Cereda M, Kayikci M, König J, Hortobágyi T, Nishimura AL, Zupunski V, Patani R, Chandran S, Rot G, Zupan B, Shaw CE, Ule J: **Characterizing the RNA targets and position-dependent splicing regulation by TDP-43.** *Nat Neurosci* 2011, **14**:452-458.
  51. Kuo PH, Chiang CH, Wang YT, Doudeva LG, Yuan HS: **The crystal structure of TDP-43 RRM1-DNA complex reveals the specific recognition for UG- and TG-rich nucleic acids.** *Nucleic Acids Res* 2014, **42**:4712-4722.
- This report reveals a clear relationship between TDP-43 and its binding to RNA.
52. Lukavsky PJ, Daujotyte D, Tollervey JR, Ule J, Stuaní C, Buratti E, Baralle FE, Damberger FF: **Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43.** *Nat Struct Mol Biol* 2013, **20**:1443-1449.
  53. Yu Z, Fan D, Gui B, Shi L, Xuan C, Shan L, Wang Q, Shang Y, Wang Y: **TDP-43 interacts with fragile X mental retardation protein (FMRP)/Staufen (STAU1) and regulates SIRT1 expression in neuronal cells.** *J Biol Chem* 2012, **287**:22560-22572.
  54. Kino Y, Washizu C, Oma Y, Onishi H, Nezu Y, Sasagawa N, Nukina N, Ishiura S: **MBNL and CELF proteins regulate alternative splicing of the skeletal muscle chloride channel CLCN1.** *Nucleic Acids Res* 2009, **37**:6477-6490.
  55. Yadav AR, Mace CR, Miller BL: **Examining the interactions of the splicing factor MBNL1 with target RNA sequences via a label-free, multiplex method.** *Anal Chem* 2014, **86**:1067-1075.
  56. Warf MB, Berglund JA: **MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T.** *RNA* 2007, **13**:2238-2251.
  57. Ayala YM, De Conti L, Avendaño-Vázquez SE, Dhir A, Romano M, D'Ambrogio A, Tollervey J, Ule J, Baralle M, Buratti E, Baralle FE: **TDP-43 regulates its mRNA levels through a negative feedback loop.** *EMBO J* 2011, **30**:277-288.
  58. Buratti E, Brindisi A, Giombi M, Tisminetzky S, Ayala YM: **TDP-43 binds heterogeneous nuclear ribonucleoprotein A/B through its C-terminal tail, an important region for the inhibition of cystic fibrosis transmembrane conductance regulator exon 9 splicing.** *J Biol Chem* 2005, **280**:37572-37584.
  59. Zu T, Gibbens B, Doty NS, Gomes-Pereira M, Huguet A, Stone MD, Margolis J, Peterson M, Markowski TW, Ingram MA et al.: **Non-ATG-initiated translation directed by microsatellite expansions.** *Proc Natl Acad Sci U S A* 2011, **108**:260-265.
- Discovery of novel form of translation in TNR repeats.
60. Wang YT, Kuo PH, Chiang CH, Liang JR, Chen YR, Wang S, Shen JC, Yuan HS: **The truncated C-terminal RNA recognition motif of TDP-43 protein plays a key role in forming proteinaceous aggregates.** *J Biol Chem* 2013, **288**:9049-9057.
  61. Donmez G, Arun A, Chung CY, McLean PJ, Lindquist S, Guarente L: **SIRT1 protects against  $\alpha$ -synuclein aggregation by activating molecular chaperones.** *J Neurosci* 2012, **32**:124-132.
  62. Xu W, Yang H, Liu Y, Yang Y, Wang P, Kim SH, Ito S, Yang C, Wang P, Xiao MT, Liu LX, Jiang WQ, Liu J, Zhang JY, Wang B, Frye S, Zhang Y, Xu YH, Lei QY, Guan KL, Zhao SM, Xiong Y:

- Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of  $\alpha$ -ketoglutarate-dependent dioxygenases.** *Cancer Cell* 2011, **19**:17-30.
63. Chen Z, Zang J, Whetstine J, Hong X, Davrazou F, Kutateladze TG, Simpson M, Mao Q, Pan CH, Dai S, Hagman J, Hansen K, Shi Y, Zhang G: **Structural insights into histone demethylation by JMJD2 family members.** *Cell* 2006, **125**:691-702.
64. Kohli RM, Zhang Y: **TET enzymes, TDG and the dynamics of DNA demethylation.** *Nature* 2013, **502**:472-479.  
Excellent review of methylation cycle in DNA.